Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks

Mäyrä, Janne

2021-04

Mäyrä , J , Keski-Saari , S , Kivinen , S , Tanhuanpää , T , Hurskainen , P , Kullberg , P , Poikolainen , L , Viinikka , A , Tuominen , S , Kumpula , T & Vihervaara , P 2021 , ' Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks ' , Remote Sensing of Environment , vol. 256 , 112322 . https://doi.org/10.1016/j.rse.2021.112322

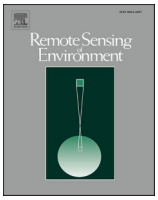
http://hdl.handle.net/10138/328801
https://doi.org/10.1016/j.rse.2021.112322

Remote Sensing of Environment 256 (2021) 112322

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse



Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks

Janne Mäyrä [a,*], Sarita Keski-Saari [b,c], Sonja Kivinen [a,c], Topi Tanhuanpää [c,d], Pekka Hurskainen [a,e], Peter Kullberg [a], Laura Poikolainen [c], Arto Viinikka [f], Sakari Tuominen [g], Timo Kumpula [c], Petteri Vihervaara [a]

[a] Finnish Environment Institute (SYKE), Biodiversity Centre, Latokartanonkaari 11, Helsinki FI-00790, Finland
[b] Department of Environmental and Biological Sciences, University of Eastern Finland, P.O. Box 111, Joensuu FI-80101, Finland
[c] Department of Geographical and Historical Studies, University of Eastern Finland, Yliopistonkatu 7, Joensuu FI-80101, Finland
[d] Department of Forest Sciences, University of Helsinki, Helsinki FI-00014, Finland
[e] Earth Change Observation Laboratory, Department of Geosciences and Geography, University of Helsinki, P.O. Box 64, Helsinki 00014, Finland
[f] Finnish Environment Institute (SYKE), Environmental Policy Centre, Latokartanonkaari 11, Helsinki FI-00790, Finland
[g] Natural Resources Institute Finland (Luke), Latokartanonkaari 9, Helsinki FI-00790, Finland

ARTICLE INFO

Keywords:
Hyperspectral imaging
Deep learning
Convolutional neural network
Tree species classification

ABSTRACT

During the last two decades, forest monitoring and inventory systems have moved from field surveys to remote sensing-based methods. These methods tend to focus on economically significant components of forests, thus leaving out many factors vital for forest biodiversity, such as the occurrence of species with low economical but high ecological values. Airborne hyperspectral imagery has shown significant potential for tree species classification, but the most common analysis methods, such as random forest and support vector machines, require manual feature engineering in order to utilize both spatial and spectral features, whereas deep learning methods are able to extract these features from the raw data.

Our research focused on the classification of the major tree species Scots pine, Norway spruce and birch, together with an ecologically valuable keystone species, European aspen, which has a sparse and scattered occurrence in boreal forests. We compared the performance of three-dimensional convolutional neural networks (3D-CNNs) with the support vector machine, random forest, gradient boosting machine and artificial neural network in individual tree species classification from hyperspectral data with high spatial and spectral resolution. We collected hyperspectral and LiDAR data along with extensive ground reference data measurements of tree species from the 83 km^2 study area located in the southern boreal zone in Finland. A LiDAR-derived canopy height model was used to match ground reference data to aerial imagery. The best performing 3D-CNN, utilizing 4 m image patches, was able to achieve an F1-score of 0.91 for aspen, an overall F1-score of 0.86 and an overall accuracy of 87%, while the lowest performing 3D-CNN utilizing 10 m image patches achieved an F1-score of 0.83 and an accuracy of 85%. In comparison, the support-vector machine achieved an F1-score of 0.82 and an accuracy of 82.4% and the artificial neural network achieved an F1-score of 0.82 and an accuracy of 81.7%. Compared to the reference models, 3D-CNNs were more efficient in distinguishing coniferous species from each other, with a concurrent high accuracy for aspen classification.

Deep neural networks, being black box models, hide the information about how they reach their decision. We used both occlusion and saliency maps to interpret our models. Finally, we used the best performing 3D-CNN to

* Corresponding author.
E-mail addresses: janne.mayra@syke.fi (J. Mäyrä), sarita.keski-saari@uef.fi (S. Keski-Saari), sonja.i.kivinen@syke.fi (S. Kivinen), topi.tanhuanpaa@helsinki.fi (T. Tanhuanpää), pekka.hurskainen@syke.fi (P. Hurskainen), peter.kullberg@syke.fi (P. Kullberg), laura.poikolainen@uef.fi (L. Poikolainen), arto.viinikka@syke.fi (A. Viinikka), sakari.tuominen@luke.fi (S. Tuominen), timo.kumpula@uef.fi (T. Kumpula), petteri.vihervaara@syke.fi (P. Vihervaara).

https://doi.org/10.1016/j.rse.2021.112322
Received 1 June 2020; Received in revised form 28 January 2021; Accepted 29 January 2021
Available online 12 February 2021

produce a wall-to-wall tree species map for the full study area that can later be used as a reference prediction in, for instance, tree species mapping from multispectral satellite images. The improved tree species classification demonstrated by our study can benefit both sustainable forestry and biodiversity conservation.

1. Introduction

Recent advances in remote sensing technology hold much promise for the detailed mapping of the spatiotemporal distribution and characteristics of tree species over wide areas (Fassnacht et al., 2016). Some of the most promising techniques related to tree species classification and single tree detection are based on hyperspectral and light detection and ranging (LiDAR) data, particularly in boreal and temperate ecosystems (e.g., Jones et al. (2010); Maschler et al. (2018); Roth et al. (2015a, 2015b)), even though there are still unsolved challenges especially in tropical ecosystems (Asner and Martin, 2009; Baldeck et al., 2015). Hyperspectral images include a contiguous spectral range with hundreds of narrow bands. Hyperspectral sensors typically operate on visible and near-infrared (VNIR) area of the electromagnetic spectrum (400–1100 nm), and modern sensors are also able to utilize the shortwave infrared (SWIR) area of the spectrum (1100–2500 nm). In contrast to traditional optical imagery with up to four bands (red, green, blue and near-infrared (NIR)) or multispectral imagery with up to twenty bands, this rich spectral range can be used to distinguish minor differences in the spectral signatures of different materials (Goetz et al., 1985; Melgani and Bruzzone, 2004).

In addition to typical challenges with aerial imagery, such as atmospheric effects and varying illumination conditions, having numerous spectral bands leads to a complex structure and large size of the data and requires efficient analysis methods. Machine learning methods, such as support-vector machines (SVMs), random forests (RF), gradient boosting machines (GBMs) and artificial neural networks (ANN) have been used in various remote sensing tasks. In particular, SVM has been extensively used for tree species identification from airborne hyperspectral data, and was identified as the most common machine learning method for these types of tasks by Fassnacht et al. (2016). More recently, Kandare et al. (2017) and Dalponte et al. (2019) achieved promising results with SVM with overall classification accuracies of 80% for three different species and 88% for nine different species respectively. Modzelewska et al. (2020) used SVM to produce a tree species map for Białowieża Forest in Poland, showing that this method is also suitable for accurately mapping tree species across larger areas instead of only in small study sites. RF and ANN were used by Nevalainen et al. (2017) for classifying unmanned aerial vehicle imagery into four different tree species, achieving an overall accuracy of around 95%, an F1-score (the harmonic mean of the user's and the producer's accuracies) of 0.93 and a Kappa score of 0.9 with both methods.

Since the beginning of the 2010s, convolutional neural networks (CNNs) have been the de-facto approach for computer vision tasks, such as image classification, object detection and semantic segmentation. Even though CNNs were first proposed as early as the late 1980s (LeCun et al., 1989), they gained larger interest only after AlexNet (Krizhevsky et al., 2012) won the ImageNet Large Scale Visual Recognition Challenge in 2012. Since then, different CNN models have been tailored for one-dimensional input data, such as a single spectral signal (1D-CNN), two-dimensional input features such as photographs (2D-CNN) and also three-dimensional inputs such as hyperspectral cubes or volumetric data (3D-CNN) (Audebert et al., 2019; Paoletti et al., 2019).

The main advantage of deep learning methods over more traditional machine learning methods, such as SVM and RF, is that they are able to automatically extract features from input data, and that they can also utilize spatial (2D-CNN) and spectral-spatial (3D-CNN) information instead of spectral information alone, whereas traditional machine learning methods are heavily reliant on hand-crafted features. Selecting and generating these features, a process known as *feature engineering*, requires both manual work and heavy domain expertise (e.g., (e.g., Sothe et al. (2020)). For tree species classification tasks a typical feature engineering process consists of computing various vegetation indices and textural features. Instead, CNNs work on raw data and are automatically able to extract significant features, some of which manual feature engineering may ignore.

The majority of studies utilizing deep learning and remote sensing from hyperspectral imagery are focused on land use and land cover (LULC) classification tasks due to the most common hyperspectral benchmark datasets (Indian Pines, Pavia and Salinas) being LULC tasks (Audebert et al., 2019; Ma et al., 2019; Paoletti et al., 2019). Out of the studies focusing on tree species identification, Sothe et al. (2020) compared the performance of 2D-CNN with SVM and RF for tree-species classification in Southern Brazilian forests with 14 target species; their CNN implementation outperformed others with overall accuracies of 84.4% and 74.95% for two different study areas. For boreal forests, Trier et al. (2018) compared the effectiveness of partial least squares regression, pixel classification based on conifer and spruce indices, a 2D-CNN and a 1D-CNN for the classification of boreal forest tree species into three target species from hyperspectral data. Out of these, the 1D-CNN achieved an 87% accuracy, while the 2D-CNN (74% accuracy) was outperformed not only by the 1D-CNN, but also by partial least squares regression (78% accuracy). Their deep learning implementations, however, did not utilize all collected hyperspectral data, as they discarded all of the SWIR data and only used 160 bandwidths from the VNIR sensor for the 1D-CNN and three bandwidths blended with vegetation height for the 2D-CNN. Furthermore, some recent studies have utilized full spectral information and deep learning methods in tree species classification in boreal forests. For example, Pölönen et al. (2018) proposed a 3D-CNN approach for tree species classification utilizing both UAV-collected 33-band hyperspectral data and a normalized canopy height model, and achieved an overall accuracy of 96.2% with three target species. This indicates that even experimental 3D-CNN models are able to achieve either similar or better results compared to other classification methods for the data originally presented in Nevalainen et al. (2017).

Nowadays, forest monitoring and inventory systems based on multisource remote sensing (RS) data efficiently produce information on economically significant components of forests, i.e., the growing stock of a few main tree species (see e.g., Maltamo and Packalen (2014); Næsset (2002); Nevalainen et al. (2017); Packalén and Maltamo (2007)). From the perspective of sustainable forestry and forest biodiversity management, there is a knowledge gap concerning the occurrence of minor deciduous tree species that diversify the forest structure and have important ecosystem functions. For example, old, large-diameter aspens support high numbers of species, including numerous red-listed species (Kivinen et al., 2020; Rassi et al., 2010), and they have been included as ecologically relevant individuals in studies that aim to map aspen abundance (Latva-Karjanmaa et al., 2007; Maltamo et al., 2015; Viinikka et al., 2020). In Finland, ecologically significant components of forest structure, such as scattered deciduous tree species with low commercial value were collected along with other forest parameters in compartment-wise forest inventory (Poso, 1983). The shift to RS-based system in the early 2010s has practically ended the extensive compartment-level field measurements, and detailed information on minor deciduous tree species (e.g. European aspen, *Populus tremula* L., see Kivinen et al. (2020)) is not available, as they are pooled in one class in the system. Improved tree species detection with hyperspectral data could enable the simultaneous detection of both economically and ecologically important tree species and facilitate the consideration of

multiple values of forests.

In this study, we focus on boreal forest ecosystems to demonstrate the use of effective analysis methods such as deep learning (i.e., 3D-CNN), for the first time in our knowledge, for a big hyperspectral data set. Our study concerns three major tree species, Scots pine (*Pinus sylvestris*), Norway spruce (*Picea abies* (L.) Karst.) and birch (*Betula* sp.), as well as a keystone species, European aspen that has a scattered occurrence in boreal forests. The research aims at answering the following questions:

1. How does the 3D-CNN perform in comparison with SVM, RF, GBM and ANN in tree species classification?
2. How accurately can the four common boreal tree species be recognized from hyperspectral data at the tree level?

2. Materials

2.1. Study area

Our study area is located in the Evo forest area in Hämeenlinna, Southern Finland, and consists of southern boreal forests (Fig. 1). The 83 km^2 study area is mostly managed but also covers two important conservation areas. These conservation areas cover a total of 7 km^2. The forests are mostly dominated by Norway spruce and Scots pine with a mixture of Downy birch (*Betula pendula*) and Silver birch (*Betula pubescens*). European aspen and other deciduous species (e.g., *Larix sibirica*, *Sorbus aucuparia*, and *Alnus incana*) are rather scarce in the dominant canopy layer.

2.2. Hyperspectral and airborne laser scanning data

Hyperspectral and LiDAR data were captured on July 16th, 2018 in the morning under cloud-free conditions for the whole study area, with the sun angle varying between 27° and 44°. Data were collected from 1500 m altitude, resulting in 0.5 m spatial resolution for the VNIR data, 1.0 m spatial resolution for the SWIR data and 10.2 p/m^2 for the LiDAR data. All fo the details of airborne data collection are presented in Table 1.

The hyperspectral and LiDAR data were georeferenced and orthorectified using Parge 3.4 software (Richter and Schläpfer, 2002). Hyperspectral cubes were orthorectified based on digital surface model generated from LiDAR data using fast-nearest neighbor interpolation. The atmospheric correction for hyperspectral data was performed by a contractor with ATCOR software version 4.7.3 by ReSe (Richter and Schläpfer, 2004). This correction did not account for bidirectional reflectance distribution or shadows but rather meant to remove atmospheric artefacts in captured data. Thus, the result is not the true surface

Table 1
Flight index information.

Time of data capture	2018.07.16 08:27–11:14
VNIR camera	HySpex 1800 – SN00827
VNIR spectral range	406–995 nm, 186 bands
SWIR camera	HySpex 384 me – SN3126
SWIR spectral range	956–2525 nm, 288 bands
LiDAR scanner	Leica ALS70-HP – SN7204
LiDAR point cloud density	10.2 p/m^2
Aircraft	Piper PA-31-350 Chieftain – LN-TTC
Maximum flight altitude	1500 m above ground level
Solar angle during data acquisition	26° – 44°

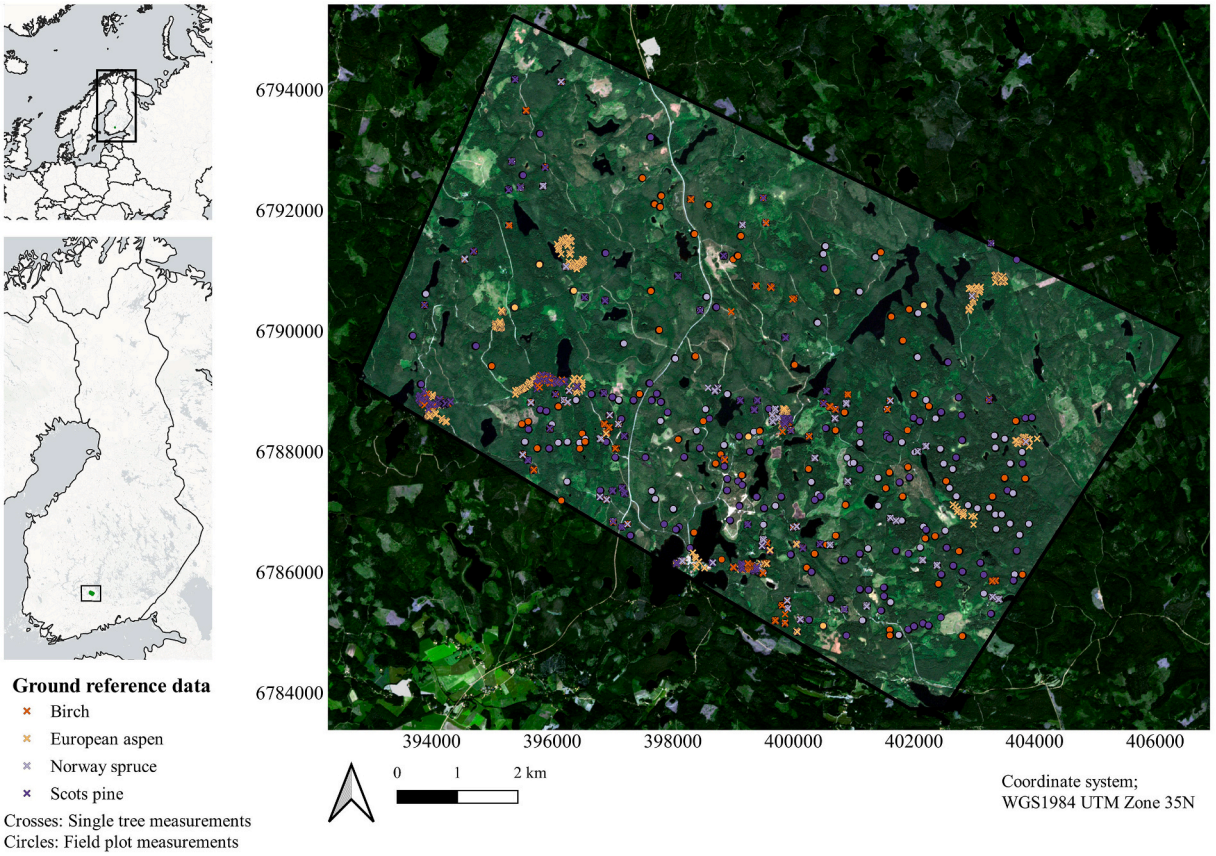


Fig. 1. Study area and ground reference data locations. Hyperspectral and LiDAR data are captured from the highlighted area. Map produced from ESA remote sensing data (Sentinel-2 imagery, bands B04, B03 and B02) captured on July 19th, 2018.

reflectance but rather an approximation. Atmospheric correction was also performed to correct spectral misregistration, which leads to a small shift in wavelengths. Spectral bands in the central wavelengths of 720 nm, 895–1000 nm, 1081–1191 nm, 1332–1469 nm and 1780–2021 nm were interpolated due to either water or oxygen absorption in the atmosphere. In addition, SWIR bands from band 274 (2449.44 nm) onward were masked out due to poor signal-to-noise levels. There was a slight overlap between the last bands of the VNIR sensor and the first bands of the SWIR sensor, but they fell under the interpolated area and were discarded in our analyses. Hyperspectral data was tiled into a total of 381 tiles of 500 × 500 m. A normalized canopy height model (CHM) with a spatial resolution of 0.5 m was derived from LiDAR data.

Because SWIR channels had a coarser spatial resolution than both the VNIR and LiDAR derived CHM, all SWIR data were upsampled to have the same spatial resolution as other data using the nearest neighbor interpolation method. After this, in order to simplify further processing, the VNIR, SWIR and CHM were concatenated to single image stacks with the spatial extent of 500 × 500 m each. The alignment of CHM, VNIR and SWIR after upsampling and stacking was inspected visually using buildings and other landmarks.

After further data exploration, bands with a central wavelength of 2000 nm onward were discarded due to issues of low quality such as water bodies with abnormally high reflectance values. In addition, we discarded all of the interpolated bands, and our final data had 250 spectral bands with central wavelengths in the ranges of 401.32–717.49, 723.72–892.05, 1006.17–1077.20, 1197.67–1329.06 and 1471.22–1776.93 nm.

2.3. Ground reference data

The ground reference data were measured during the summer of 2018 using both circular plots and individual tree measurements. Overall, 400 circular field plots, each with a 9 m radius were distributed over the study area using stratified sampling. The study area was stratified according to the main tree species (5 strata), DBH (5 strata), and basal area (4 strata) using compartment-level forest inventory data from 2015. The compartments were first considered as lists of IDs belonging to each stratum. Measured compartments were selected systematically from the lists and the number of field plots in each stratum was determined by the stratum's proportional area within the study site. Primary locations for the field plots were set in the center of each measured compartment. The final plot centers were positioned using a real-time kinematic global navigation satellite system (Topcon RTK-GNSS and Trimble RTK devices). The locations of individual trees within the plots were defined using the azimuth angle and the distance from the plot centers.

To ensure the visibility of reference trees from above, only trees with a diameter at breast height (DBH) of 150 mm or more were included in the reference data. Only pines, spruces, both birch species and aspens were included in the training and validation sets, because the number of other species in our ground reference data was small (less than 120 trees with DBH ≥ 150 mm). Also, Silver and Downy birch were combined into one class. In addition to tree observations from circular plots, the locations of individual trees were recorded from the study area during the summer and fall of 2019 using an RTK-GNSS device. In total, the reference data consisted of 4343 trees from circular plots and 2256 trees measured individually using the RTK-GNSS device. The species distributions of the ground reference data are presented in Table 2.

Table 2
Numbers of trees with DBH ≥ 150 mm by collection method.

Tree species	Field plots	Single tree	Total
Scots pine	1882	688	2570
Norway spruce	1550	495	2045
Downy and silver birch	793	474	1267
European aspen	118	599	717
Overall	4343	2256	6599

3. Methods

Our aim was to test the performance of various machine learning and deep learning models in individual tree species identification in a boreal forest using airborne hyperspectral data, with a special focus on European aspen. Our task was divided into four separate subtasks:

1. detect and segment individual trees from airborne data,
2. match detected and segmented tree crowns to ground reference data,
3. utilize these data to fit the models,
4. use the models to classify unlabeled trees.

We performed the first two steps with commonly used methods, and for tree species classification we focused on comparing the efficiency of several different techniques ranging from traditional machine learning to state-of-the-art deep learning methods. Our analyses and the source codes that we used are available at https://github.com/jaeeolma/tree-detection-evo.

3.1. Matching airborne data with ground reference data

We utilized LiDAR-derived CHM to match individual trees from ground reference data to airborne data to allow us to control the minimum height for detected trees as well as being able to segment shadowed areas. For individual tree crown delineation, we used the algorithm proposed by Dalponte and Coomes (2016), as it has been shown to perform well in a reasonable time (e.g., Liu et al. (2019)). All treetop detection and tree delineation were performed with R version 3.5 and *lidR* package version 2.2.1 (Roussel et al., 2017). First, the tiled CHMs were smoothed with a low pass filter, and then initial treetops were detected with a local maximum filter with a circular moving window, using a window size of 5 m and a minimum height of 10 m. Individual trees were segmented based on these treetops using the *dalponte2016* function from lidR with a minimum height of 10 m (`th_tree`), a growing threshold 1 of 0.65 (`th_seed`), a growing threshold 2 of 0.5 (`th_cr`) and a maximum crown diameter of 5 m (`max_cr`). The parameters for the *dalponte2016* function were selected in order to ensure the detection of trees in the upper canopy with a DBH of at least 150 mm and to avoid segmentations being mixed with the neighboring trees, ensuring that it is more likely that field data points located within a segment really correspond to that tree crown. As a post-processing step, a 2D convex hull was applied to results from previous steps in order to have convex tree crowns (Dalponte and Coomes, 2016).

After segmenting the full study area, we matched the delineated tree crowns and our field measurements with the following algorithm: For each tree crown segment, we checked whether it contained any field data measurements. If there was only one field data point within a tree crown, then the tree crown was labeled with this field data point. If two or more measured trees were located inside one tree crown segment, then we used the following rules: If any of these field data points was individually measured, we only considered individually measured trees within the segmented crown as a valid label for the corresponding tree crown due to their higher spatial accuracy. Finally, the tree crown was labeled with the closest remaining field data measurement to the detected treetop pixel.

Because extracted image patches can contain multiple labeled trees, using randomized split can lead to data leakage between training and validation sets. Using this kind of data to train our models can give overly optimistic validation results that do not reflect the models generalization power on truly unseen data, because it is likely that models have seen at least parts of the image patches used for validation (Audebert et al., 2019; Meyer et al., 2019). To address this, we split our labeled tiles into disjoint training and validation sets based on the tiling

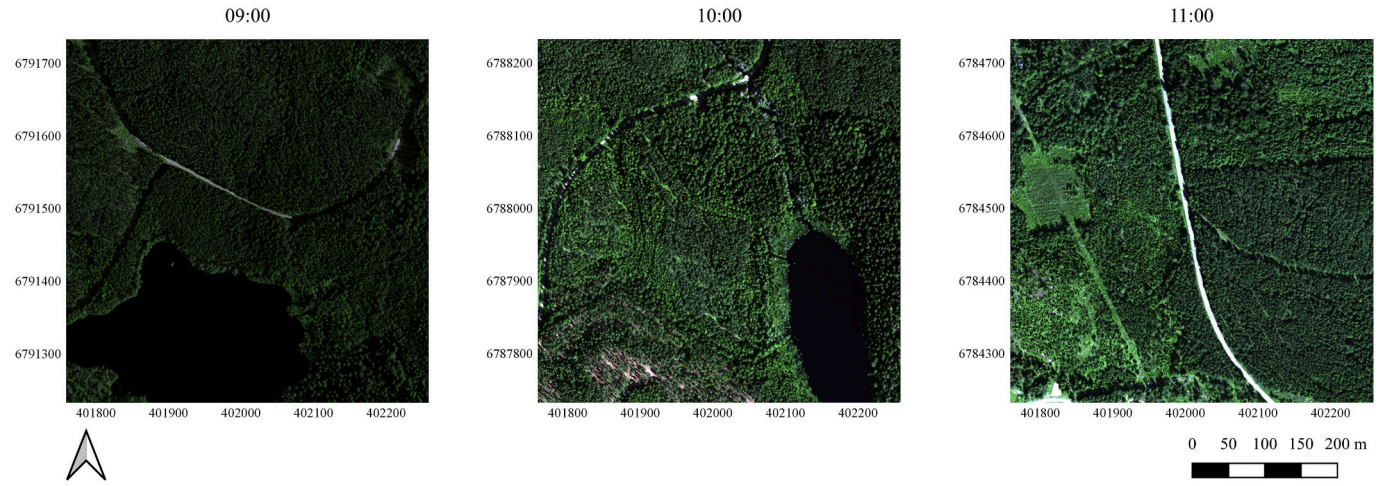


Fig. 2. Examples of different daylight illuminance conditions at approximately 09:00 (R8C19), 10:00 (R15C19) and 11:00 (R22C19). All images are composed of hyperspectral images (Red: 664 nm, Green: 560 nm, Blue: 493 nm) and colour bands are scaled with respect to R15C19. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

columns, using columns 19–22 as our validation set. Because flight lines of data collection were almost perpendicular to tiling column identifiers, by splitting the column identifiers we could ensure that both sets contain data collected in various time steps of data collection and thus they have all the possible illuminance conditions present in our data. An example of these differences is presented in Fig. 2.

For CNN models, we adopted the typical approach for hyperspectral remote sensing scene classification and extracted square image patches centered around a labeled treetop. The correct label for each patch is then the label for the corresponding treetop, no matter how many other trees are present in the image patch. Square patches with diameters of 4, 6, 8 and 10 m (image cubes with dimensions of $250 \times 9 \times 9$, $250 \times 13 \times 13$, $250 \times 17 \times 17$ and $250 \times 21 \times 21$ pixels, respectively) were extracted in order to test whether the image patch size would provide differences in the classification results. Larger image patches will contain multiple trees from multiple species and this information could be useful for classification purposes. For example, pines tend to be surrounded by other pines. We did not use any information about delineated tree crowns in CNN classifications and only used the segmented results to produce the final wall-to-wall maps.

For our reference methods, we computed the summary statistics (mean and standard deviation) of the full spectra for each delineated tree crown. This totaled 500 features for each tree crown object.

For all methods except RF and GBM, the input reflectance values were normalized with the mean and standard deviation of the training set such that each spectral channel in the training set has zero mean and unit variance. This is a standard preprocessing step for these models, as it both prevents features with large values from dominating the classification process and speeds up the convergence of deep learning methods (LeCun et al., 2012). Due to the nature of how decision tree-based models (RF and GBM) are constructed, normalization doesn't affect the performance of these models at all, and thus this step could be omitted.

3.2. Classification methods

There are several ways to utilize CNNs in hyperspectral image recognition. First of all, a CNN using only one-dimensional convolutions (1D-CNN) can be used to extract features from a single pixel spectra. This method, however, completely ignores the spatial features present. Another way is to use a similar approach that is commonly used in RGB-image recognition and use two-dimensional convolution kernels (2D-CNN) that are applied to each input channel separately. The problems with this approach are that it ignores the rich spectral information, and the number of filters and thus parameters that must be optimized can be high, because their number is proportional to the dimensions of the input. For RGB-images, a simple convolutional layer with a kernel size of 3×3 and 32 output channels will have $3 \times 3 \times 32 \times 3 = 864$ parameters, but hyperspectral image cubes might have over 200 input channels. This type of data leads to at least $3 \times 3 \times 32 \times 200 = 57600$ parameters in a single layer. Because of this, 2D-CNN approaches typically perform some kind of dimensionality reduction, such as principal component analysis (PCA) or minimum noise fraction (MNF), on the input data (Audebert et al., 2019; Paoletti et al., 2019).

Recent studies have shown that CNNs utilizing both spectral and spatial information yield better results than those that use only one of these types of information (Audebert et al., 2019). It is possible to first extract spatial features with a 2D-CNN and then the spectral features with a 1D-CNN. However, this method does not help to solve the problem with the high number of parameters related to the 2D-CNN method. The alternative method that has shown the most promising results is to extract spectral and spatial features simultaneously with three-dimensional convolutions (3D-CNN). Convolutional layers in a 3D-CNN produce feature cubes instead of one-dimensional feature vectors (like 1D-CNN) or two-dimensional feature maps (2D-CNN), and they are thus able to extract features that are more complex than handcrafted features (Audebert et al., 2019; Paoletti et al., 2019).

Our CNN models are fairly simple, consisting of four or five (for 10 m image patches) convolutional layers with three-dimensional kernels, followed by two linear layers. Convolutional layers were used for extracting spectral-spatial features from input data, and linear layers performed the final classification from these features. Before the linear layers, the input is converted to a 1-dimensional vector format (*flattened*). Because our input has an odd number of pixels per spatial dimension, we did not use any pooling layers, but rather shrunk the input by not having zero padding and having a stride (the amount of pixels that the convolutional kernel is moved during one step) of 2 in some of the layers. Kernel sizes and the strides of convolutional layers are selected such that all of the input data is used, and the data has spatial dimensions of 1×1 after the final convolutional layer. We used the rectified linear unit (ReLU, $ReLU(x) = max(0, x)$) as the activation function in all convolutional layers and the first linear layer. In order to get probability values for classes from raw output values, the activation function for the final layer was softmax, defined as

$$softmax(x)_i = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}, \quad (1)$$

Table 3
Summary of CNN architecture for different image patch sizes.

	Input	4 m	6 m	8 m	10 m
		1 × 250 × 9 × 9	1 × 250 × 13 × 13	1 × 250 × 17 × 17	1 × 250 × 21 × 21
Conv1	Kernel	10 × 3 × 3	10 × 3 × 3	10 × 3 × 3	10 × 3 × 3
	Stride	2 × 1 × 1	2 × 1 × 1	2 × 1 × 1	2 × 1 × 1
	Output	32 × 121 × × 7 × 7	32 × 121 × 11 × 11	32 × 121 × 15 × 15	32 × 121 × 19 × 19
Conv2	Kernel	5 × 3 × 3	5 × 3 × 3	5 × 3 × 3	5 × 3 × 3
	Stride	2 × 1 × 1	2 × 2 × 2	2 × 2 × 2	2 × 2 × 2
	Output	64 × 59 × 5 × 5	64 × 59 × 5 × 5	64 × 59 × 7 × 7	32 × 59 × 9 × 9
Conv3	Kernel	3 × 3 × 3	3 × 3 × 3	3 × 3 × 3	3 × 3 × 3
	Stride	2 × 1 × 1	2 × 1 × 1	2 × 2 × 2	2 × 1 × 1
	Output	64 × 29 × 3 × 3	64 × 29 × 3 × 3	64 × 29 × 3 × 3	64 × 29 × 7 × 7
Conv4	Kernel	3 × 3 × 3	3 × 3 × 3	3 × 3 × 3	3 × 3 × 3
	Stride	1 × 1 × 1	1 × 1 × 1	2 × 1 × 1	1 × 2 × 2
	Output	128 × 27 × 1 × 1	128 × 27 × 1 × 1	128 × 27 × 1 × 1	64 × 27 × 3 × 3
Conv5	Kernel	Not used	Not used	Not used	3 × 3 × 3
	Stride				1 × 1 × 1
	Output				128 × 25 × 1 × 1
Linear1	Input	1 × 3456	1 × 3456	1 × 3456	1 × 3200
	Output	1 × 512	1 × 512	1 × 512	1 × 512
Linear2	Input	1 × 512	1 × 512	1 × 512	1 × 512
	Output	1 × 4	1 × 4	1 × 4	1 × 4

where x_i is the raw prediction value for class i and n is the number of classes. Implemented CNN architectures are presented in Table 3.

All CNN models implemented batch normalization (Ioffe and Szegedy, 2015) on each layer group and implemented dropout (Srivastava et al., 2014) with a probability of 0.5 between linear layers. All deep learning models are trained with the AdamW optimizer (Loshchilov and Hutter, 2019) with the parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e^{-8}$ and weight decay rate of 0.01. These values are the recommended default values tested to work well in various tasks. We also used one cycle learning rate scheduling (Smith, 2018), in which instead of having the same learning rate during the whole training process, we start with a slightly lower learning rate and periodically increase it during the first 30% of batches. After reaching the maximum, the learning rate is slowly annealed until the end. We used the maximum learning rate of 0.001, a batch size of 64 and trained for a maximum of 100 epochs, saving the model with the best validation F1-score. Training data augmentation and other regularization techniques are described in Section 3.3.

CNN models were compared with other widely used machine learning classification methods for remote sensing applications: random forest (RF), a support vector machine (SVM), a gradient boosting machine (GBM) and a feedforward neural network with two hidden layers (ANN). Optimal hyperparameters for SVM, RF and GBM were searched using randomized search (Bergstra and Bengio, 2012) on 15 parameter combinations with five-fold random cross-validation on training data, and then the model with the best cross-validation score was fitted with the full training data. For deep learning models, the iteration with the highest validation accuracy was saved and used for inference instead of the model from the last epoch. All evaluations are done with the validation set, which is otherwise not used for fitting models.

Deep learning models (CNN and ANN) were implemented with *PyTorch* version 1.4 (Paszke et al., 2017) and *fastai2* version 0.16 (Howard and Gugger, 2020), using NVIDIA V100 GPGPU. The SVM and RF models were implemented with *scikit-learn* version 0.22.1 (Pedregosa et al., 2011), whereas GBM was implemented with *LightGBM* version 2.3.0 (Ke et al., 2017). All data processing and model training was done using the computation nodes of Puhti supercomputer hosted by the CSC – IT Center for Science, Finland (CSC – IT Center for Science Finland, 2020).

3.3. Data augmentation for CNNs

As mentioned earlier, one of the problems with common hyperspectral datasets is the low number of labeled training samples. For instance, state-of-the-art image recognition models suited for RGB images have been pretrained with the ImageNet dataset (Deng et al., 2009), which has about 1.3 million annotated samples from 1000 different classes, while the most common hyperspectral benchmark datasets have less than 60,000 labeled items (usually pixels). Our data has only around 3000 labeled samples, which can be considered to be a small dataset for training deep learning models from scratch. In order to prevent overfitting and make our models able to classify unseen data more accurately, we regularized the learning process using modern techniques. We used the techniques described in this section only for CNN models, not on comparison methods. Also, no augmentations were performed in advance, but rather on-the-fly when drawing samples into a minibatch.

First of all, in order to practically eightfold our training data, each sample was randomly rotated 90 degrees clockwise or counter-clockwise and flipped horizontally or vertically during training. Also, to account for possible different lighting conditions for the same species, reflectance values were augmented with the probability of 0.5 with the following formula:

$$x_{aug} = \sigma(logit(x) + logit(change - 0.5)),\ \sigma(x) = \frac{e^x}{e^x + 1},\ logit(x) = -log\left(\frac{1}{x} - 1\right) \quad (2)$$

where x is the original image with reflectance values scaled between 0 and 1, $\sigma(x)$ is the *sigmoid* function, *logit*(x) the logit function and *change* is a uniformly drawn number from the interval [0.8, 1.2].

In addition, we used a novel data augmentation technique: mixup. In this approach, instead of using raw images as our training data, mixup augmentation generates a linear combination of two distinct images. For example, our synthetic input image might consist of 80% spruce and 20% European aspen, and thus the correct output vector would be [0.0,0.2,0.8,0.0]. Mixup has been shown to improve classification results for image classification and speech recognition as well as add robustness in case of corrupt labels (Zhang et al., 2018).

The choice of loss function was also an important factor to consider. The most commonly used loss function for multi-class classification problems is the categorical cross-entropy, also known as the log-loss, defined as

$$\sum_{i=1}^{n} -y_i log(p_i) \quad (3)$$

where n is the number of classes, p_i is the probability of class i and y_i is 1 for the correct class and 0 for others. Minimizing this loss is equivalent to maximizing the log-likelihood of the correct label. However, this can cause the model to overfit. Because minimizing cross-entropy encourages the model to assign the full probability to the correct class, the model is not guaranteed to be able to generalize. In order to avoid this, the method suggested by Szegedy et al. (2016) was used. It presents a variation of cross-entropy loss called label smoothing cross-entropy loss, which penalizes the model for overconfidence. Instead of computing the loss with the true targets y, they are replaced with the modified targets

$$y^* = y_i(1 - \alpha) + \frac{\alpha}{n} \quad (4)$$

In our study, 0.1 was used as the value for α. While regularizing the model in this way might seem counter-intuitive, it has been shown to improve robustness at least for RGB-image classification tasks (Müller et al., 2019; Szegedy et al., 2016).

3.4. Evaluation metrics

To evaluate the performance of each method, we used the following

metrics: the overall accuracy (OA) of predictions and both macro and weighted averages of the user's accuracy (UA, also known as precision), producer's accuracy (PA, also known as recall) and F1-score. The macro averages for multi-class classification results are raw averages of class-specific metrics, while weighted averages take the support of classes into account. UA measures how many of the positive predictions were relevant, PA tells us how many of the positive results were correctly classified and F1 is the harmonic mean of UA and PA. These metrics are calculated from true positives (TP), true negatives (TN), false negatives (FN) and a total number of items (N) in the following way:

$$OA = \frac{TP + TN}{N}, UA = \frac{TP}{TP + FP}, PA = \frac{TP}{TP + FN}, F1 = \frac{2 \cdot UA \cdot PA}{UA + PA} \quad (5)$$

In addition to validating the CNN's performance with only the validation set, we also classified all of the detected trees in our study area in order to see whether the distribution of our predictions is realistic. For this task, if a tree was located near the edge of a tile such that extracting a square image patch is not possible, missing reflectance values are filled by mirroring the previous values to acquire square patches.

3.5. Model interpretation

One of the disadvantages of deep neural networks is that due to their complex structure they are considered to be *black box* models. However, there are a few techniques to gain some information on which features the model considers to be the most important. First of all, by occluding parts of the input data and inspecting how the prediction probabilities change, it is possible to interpret how the model makes its decisions to some degree (Zeiler and Fergus, 2014). Another possibility is to use the model to classify an image and then compute the gradient of the maximum predicted class with respect to the input image. Higher magnitudes of gradient signify which pixels need to be changed the least to influence the classification score the most, and they can be considered to be the most influential for the classification process (Simonyan et al., 2014). The gradient can be computed with either *vanilla backpropagation* or *guided backpropagation*. The difference between these methods is that in the guided backpropagation approach, when propagating through ReLU-layers, all negative values are masked with zero, thus guiding the results to better visualize the features that have a positive impact on class scores (Springenberg et al., 2015). Visualizations of these results are called *saliency maps*.

Most of the work in CNN interpretability is focused on RGB-imagery. However, there are a couple of studies applying this work to hyperspectral data. Pölönen et al. (2018) and Nagasubramanian et al. (2019) applied vanilla backpropagation to their models and computed magnitudes of the gradients in both spatial and spectral dimensions. In this work, we used both the occlusion method and the average magnitude of gradients acquired with vanilla backpropagation to produce saliency maps for each class separately. We occluded input images with random noise generated from uniform distribution both one spatial pixel at a time and in the spectral dimension one spectral band at a time. At each step, the change of confidence for the initially predicted class was recorded. Methods for computing gradients and saliency visualizations were adapted from the PyTorch CNN Visualizations repository (Ozbulak, 2019). However, because our input data had small spatial dimensions and were centered around the detected treetop pixel, checking spatial importance was more of a sanity check for the model rather than an accurate interpretation of which spatial locations were vital for decisions.

4. Results

4.1. Field and airborne data matching

We were able to match 2874 segments with the field data. The majority of these, a total of 2176, contained only a single field measured tree, from which 1066 were measured individually using RTK-GNSS and 1110 originated from field plots. The remaining 698 segments contained multiple field measurements of trees with DBH $\geq$ 150 mm. Of these, 500 segments contained only one tree species while 198 segments contained two or more different species. The most common species combinations in the cases of multiple species in one segment were spruce and birch (78 occasions), spruce and pine (50 occasions) and birch and pine (30 occasions). Out of 6716 field data measurements (including "Other" species), 2928 (43.6%) were not located within any delineated tree crown and thus were excluded from further analyses.

In addition to excluding all trees labeled "Other", we had to omit all trees from one of the field plots due to it being located just outside of the hyperspectral data collection area, bringing the total number of trees to 2826. Overall, deciduous species had a higher matching rate compared to coniferous trees. European aspen had the highest matching rate of 61.8%, while Norway spruce had the lowest (35.3%). Individually measured trees resulted in higher matching rates compared to field plot measurements (Table 4), with overall matching rates of 63.6% and 32.0% respectively. The matching rates were calculated by dividing the number of labeled crowns by the number of labeled field data points.

Visual inspection of the segmentation results revealed inconsistencies in the spatial accuracy between hyperspectral images and the CHM. Some of the tree segments were not aligned with the hyperspectral image tree crowns (Fig. 3). These types of misaligments occurred mostly in the edges and overlapping areas of flight lines.

On average, matched trees had a slightly larger DBH compared to all of the measured trees in the ground reference data (Fig. 4), with the exception of aspen. However, almost all of the matched aspen were individually measured, with only 19 matched trees measured from field plots. Also, individually measured trees had on average a larger DBH compared to trees from field plots. On average, aspen had the largest DBH no matter the collection method for full field data and also for matched trees. Interestingly, even though both pine and spruce had on average larger DBH than birch, both species had a lower detection rate compared to birch.

The average within-segment reflectance spectra (Fig. 5) showed that deciduous species had higher reflectance values than coniferous species, especially between the 720 nm and 1400 nm wavelengths. European aspen had overall the highest reflectance and Norway spruce had the lowest. The difference in reflectances between aspens and birches in red-edge and NIR (660–900 nm) was larger than between spruce and pine, and vice versa in the SWIR-portion of the data (1000 $\leq$ nm). However, due to the varying daylight illuminance conditions during airborne data acquisition, there was high variance within each species. The normalized data showed a large difference in reflectance between the deciduous and coniferous species, particularly in the range above 700 nm.

The tree species distributions of our training and validation sets are presented in Table 5. The numbers vary slightly between different image patch sizes due to trees occurring on a tile border where it was not possible to extract a square patch.

Table 4
Matching rates for different species.

Species	Field plots		Single tree		Total	
	Matched trees	Rate	Matched trees	Rate	Matched trees	Rate
Scots pine	650	34.5%	449	65.3%	1099	42.8%
Norway spruce	463	29.9%	258	52.1%	721	35.3%
Downy and silver birch	259	32.7%	304	64.1%	563	44.4%
European aspen	19	16.1%	424	70.8%	443	61.8%
Overall	1391	32.0%	1435	63.6%	2826	42.8%

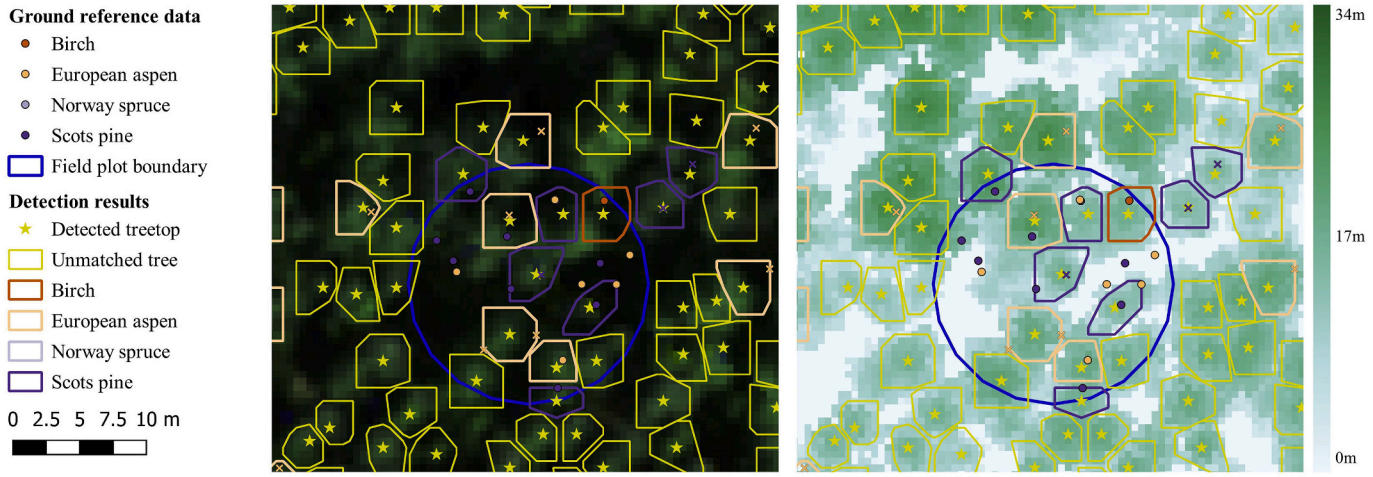


Fig. 3. Example comparison of field plot data and segmentation results for a single field plot and its surroundings. Crosses note individually measured trees and circles represent field plot measurements. Left: RGB composite from hyperspectral data, with central wavelengths of Red: 664 nm, Green: 560 nm, Blue: 493 nm. Right: LiDAR-derived canopy height model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

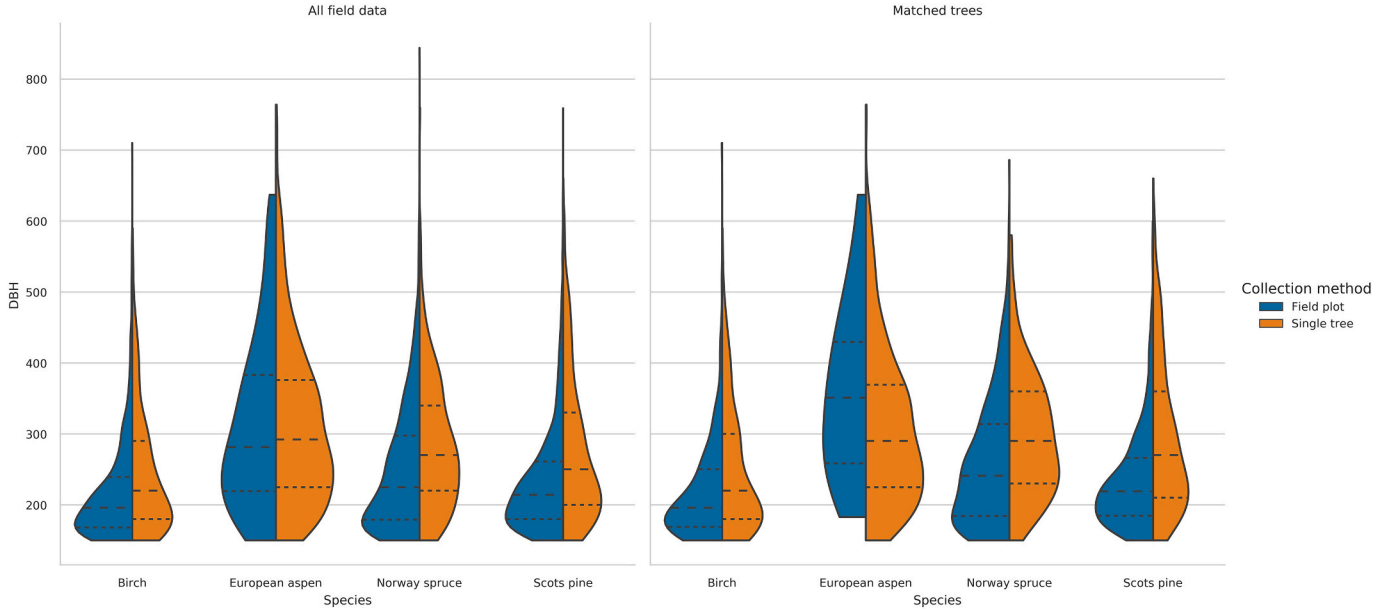


Fig. 4. Comparison of species-wise DBH-distributions for different collection methods between all data and matched data. Dashed lines within the plots show 25%, 50% and 75% quartiles for each species-method combination.

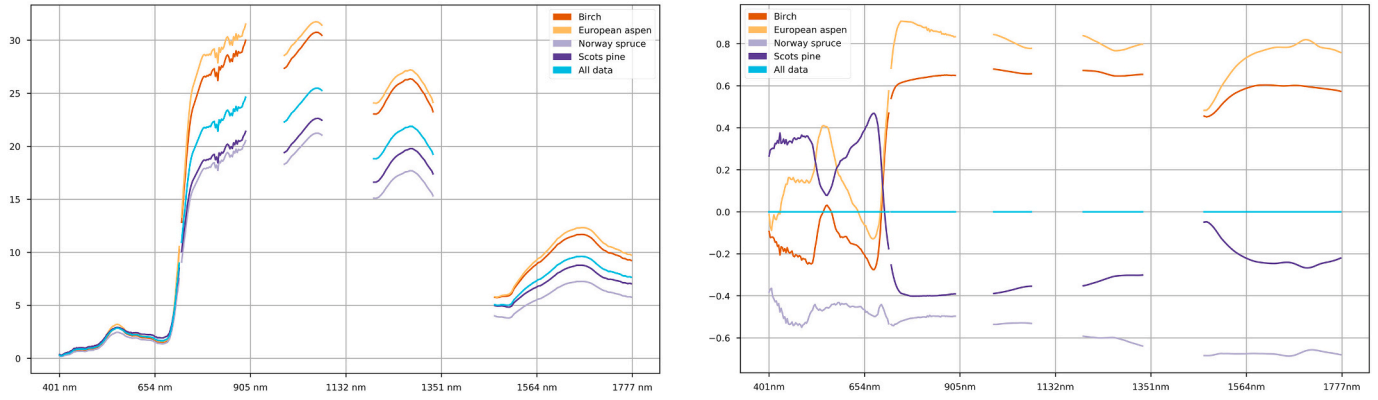


Fig. 5. Left: Average within-segment reflectances. Right: Average normalized within-segment reflectances.

Table 6
Validation set results for comparison methods.

Species	RF			SVM			LightGBM			ANN		
	UA	PA	F1	UA	PA	F1	UA	PA	F1	UA	PA	F1
Pine	0.76	0.80	0.78	0.86	0.84	0.85	0.72	0.81	0.76	0.78	0.89	0.83
Spruce	0.79	0.75	0.77	0.83	0.81	0.82	0.79	0.73	0.76	0.89	0.73	0.80
Birch	0.45	0.76	0.57	0.73	0.72	0.72	0.44	0.69	0.54	0.69	0.81	0.74
Aspen	0.80	0.34	0.48	0.83	0.93	0.87	0.76	0.32	0.45	0.88	0.89	0.88
OA	70.3%			82.4%			68.8%			81.7%		
Macro avg	0.70	0.66	0.65	0.82	0.82	0.82	0.68	0.64	0.63	0.81	0.83	0.82
Weighted avg	0.74	0.70	0.70	0.82	0.82	0.82	0.72	0.69	0.68	0.83	0.82	0.82

Table 5
Training and validation set numbers.

Species	Train	Validation
Scots pine	929	170
Norway spruce	513	208
Birch	488	75
European aspen	361	82
Overall	2291	535

4.2. Classification results

Of our reference methods, both SVM with linear kernel and ANN with two hidden layers clearly outperformed decision tree-based methods, as seen in Table 6 and Fig. 6. Decision tree-based methods had significant difficulties separating deciduous species from each other, whereas ANN and SVM only had minor classification errors. Of these methods, ANN had the highest F1-score for aspen, and SVM was the best method for pine classification. Overall, the performances of SVM and ANN were almost equal, and the selection between these methods is up to the users' preferences. However, both of these methods were outperformed by each CNN model.

The CNN using 4 m image patches had the best overall performance of all methods, beating the second best model using 6 m patches by 0.5 percentage points for OA and 0.01 for the macro F1-score. However, the size of the image patch did not have major impacts on classification accuracy, and all CNN models achieved better performance than reference methods (Table 7). All CNNs had similar results for different species. Typically, aspen and pine were classified with the highest accuracies, and birch was the most difficult species to classify correctly. The model using 4 m image patches had the best results overall due to having the highest PA (0.84) for birch and robust results for other species. Confusion matrices of all 3D-CNNs are shown in Fig. 7.

Compared to ANN and SVM, each 3D-CNN was more accurate especially in classifying the coniferous species. The most typical errors for each of the three best performing models (3D-CNN, SVM and ANN) were to incorrectly classify spruce as pine and spruce as birch. Neural network models (ANN and CNN) rarely labeled pines as spruces, and they were slightly more accurate with birch. The most accurate model for aspen classification, based on F1-score, was the CNN with 4 m image patches. However, both SVM and ANN had higher PA for aspen than any 3D-CNN.

4.3. Model interpretation

Based on the changes in prediction confidences after occlusion (Fig. 8), our model put heavy emphasis on SWIR-wavelengths between 1646 and 1700 nm, especially for coniferous species. For instance, on average the confidence for spruce classification drops more than 0.5 when input wavelengths in this range are replaced with random noise. At the same time, the confidence for pine predictions increased. Similar effects were also detected for aspen and birch; for instance, the wavelength range near 630 nm was important for positive aspen predictions (prediction confidence decreased) but had little to no effect on positive birch predictions (prediction confidence increased). Average gradient magnitudes were also higher in the wavelength ranges with the largest prediction confidence changes, thus confirming which wavelengths were the most important to the model.

According to the occlusion method, the most important spatial locations were near the center of the image (Fig. 9). Aspen was the most sensitive species for changes in the expected tree crown area, with the average classification confidence dropping as much as 0.2 in the detected treetop location. Interestingly, on average the confidence for birch predictions remained the same or even increased very little when occluding parts of the spatial input. Overall, the effects of spatial occlusion were smaller than those of spectral occlusion. Based on these results along with spatial saliency maps, we were able to confirm that the model put more importance on the area near the treetop.

4.4. Full study area classifications

We used the best performing CNN model to generate wall-to-wall tree species map for our study area. It is worth noting that our treetop data contained only trees higher than 10 m, and all trees were classified to be one of our training species. In reality, there were some situations that our methods either fail to classify at all, such as undergrowth, seedling and sapling stands, as well as some less common and rare species which were incorrectly classified as one of our four classes. The distribution of predicted species is presented in Table 8. Almost half of the detected trees in the area were classified as Scots pine, and only around 1.4% of trees were classified as European aspen.

A full tree species map is presented in Fig. 10, aggregated to 10 m spatial resolution. Each pixel was labeled with the most abundant species based on the number of treetops within the pixel. European aspen was scattered around the study area, with only a couple of larger aspen stands.

5. Discussion and future work

In this study, we compared the performance of five common machine learning methods (RF, SVM, GBM, ANN, and CNN) in identifying four main tree species in a boreal forest using airborne hyperspectral and LiDAR data. The results show that the CNN outperformed all other methods in overall accuracy. 3D-CNN models performed especially well in separating the coniferous species pine and spruce providing a beneficial method for forest industry, since commercial interests focus on conifers. Perhaps the most surprising result was that the SVM, ANN and CNNs distinguished between birch and aspen without difficulty, but had more errors with classifying spruce as birch. Spruce was also often misclassified as pine. Pine and spruce have been considered separate classes in species classifications based on hyperspectral data, whereas deciduous trees have been combined into a single class (Dalponte et al., 2014; Dalponte and Coomes, 2016). It has been suggested that young spruce trees may resemble mature birches in their spectrum due to being brighter than mature spruce trees (Trier et al., 2018). The high accuracy

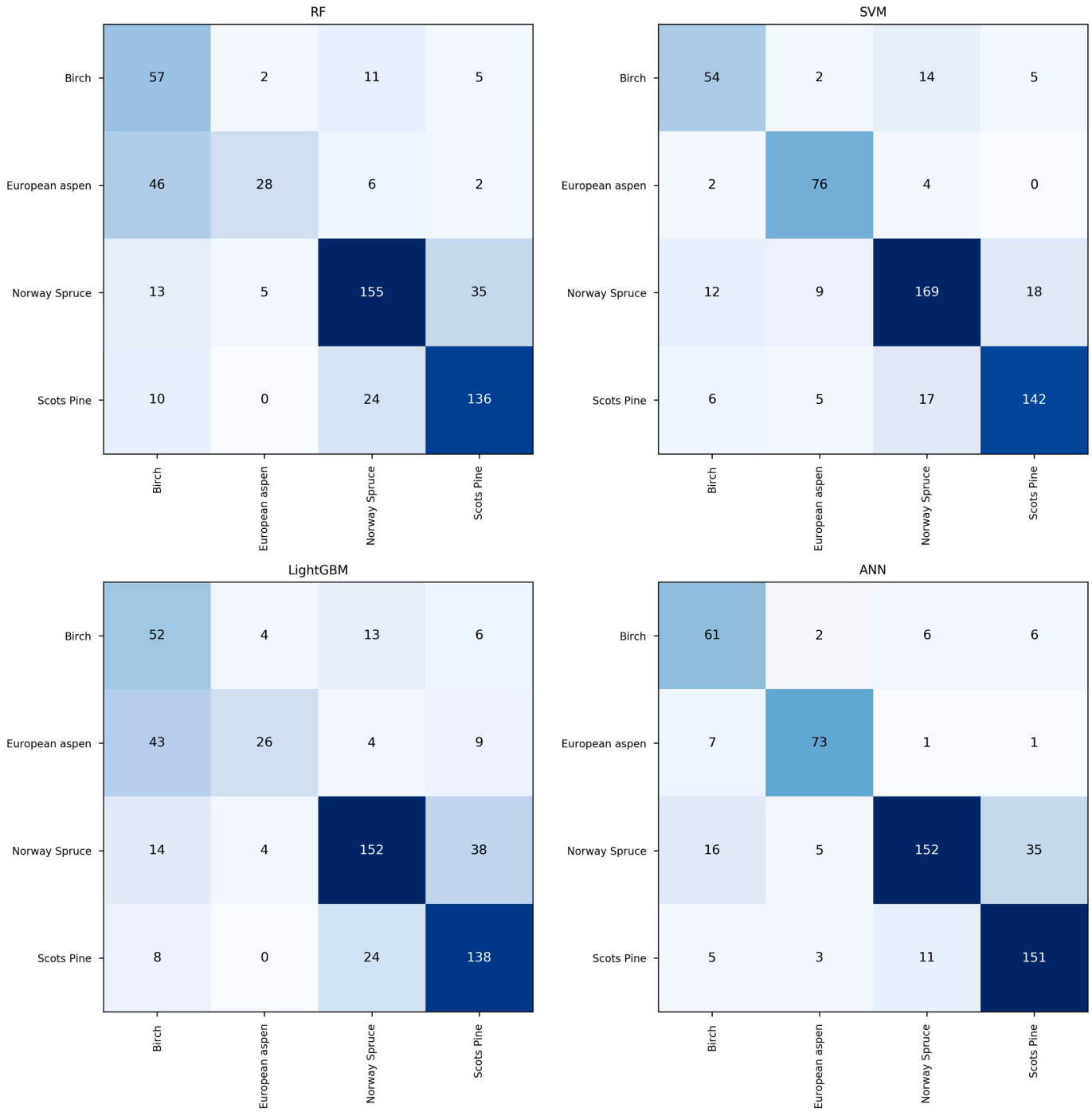


Fig. 6. Confusion matrices for comparison methods. Rows indicate correct labels and columns indicate predicted labels.

Table 7
Validation set results the CNN models.

Species	4 m			6 m			8 m			10 m		
	UA	PA	F1	UA	PA	F1	UA	PA	F1	UA	PA	F1
Pine	0.87	0.93	0.90	0.87	0.95	0.91	0.84	0.93	0.88	0.86	0.93	0.89
Spruce	0.92	0.83	0.87	0.90	0.85	0.88	0.87	0.83	0.85	0.88	0.85	0.86
Birch	0.71	0.84	0.77	0.71	0.82	0.76	0.66	0.75	0.73	0.74	0.71	0.70
Aspen	0.94	0.88	0.91	0.93	0.78	0.85	0.92	0.85	0.89	0.92	0.80	0.86
OA	87.0%			86.5%			85.1%			85.0%		
Macro avg	0.86	0.87	0.86	0.85	0.85	0.85	0.85	0.84	0.84	0.84	0.82	0.83
Weighted avg	0.88	0.87	0.87	0.87	0.87	0.87	0.85	0.85	0.85	0.85	0.85	0.85

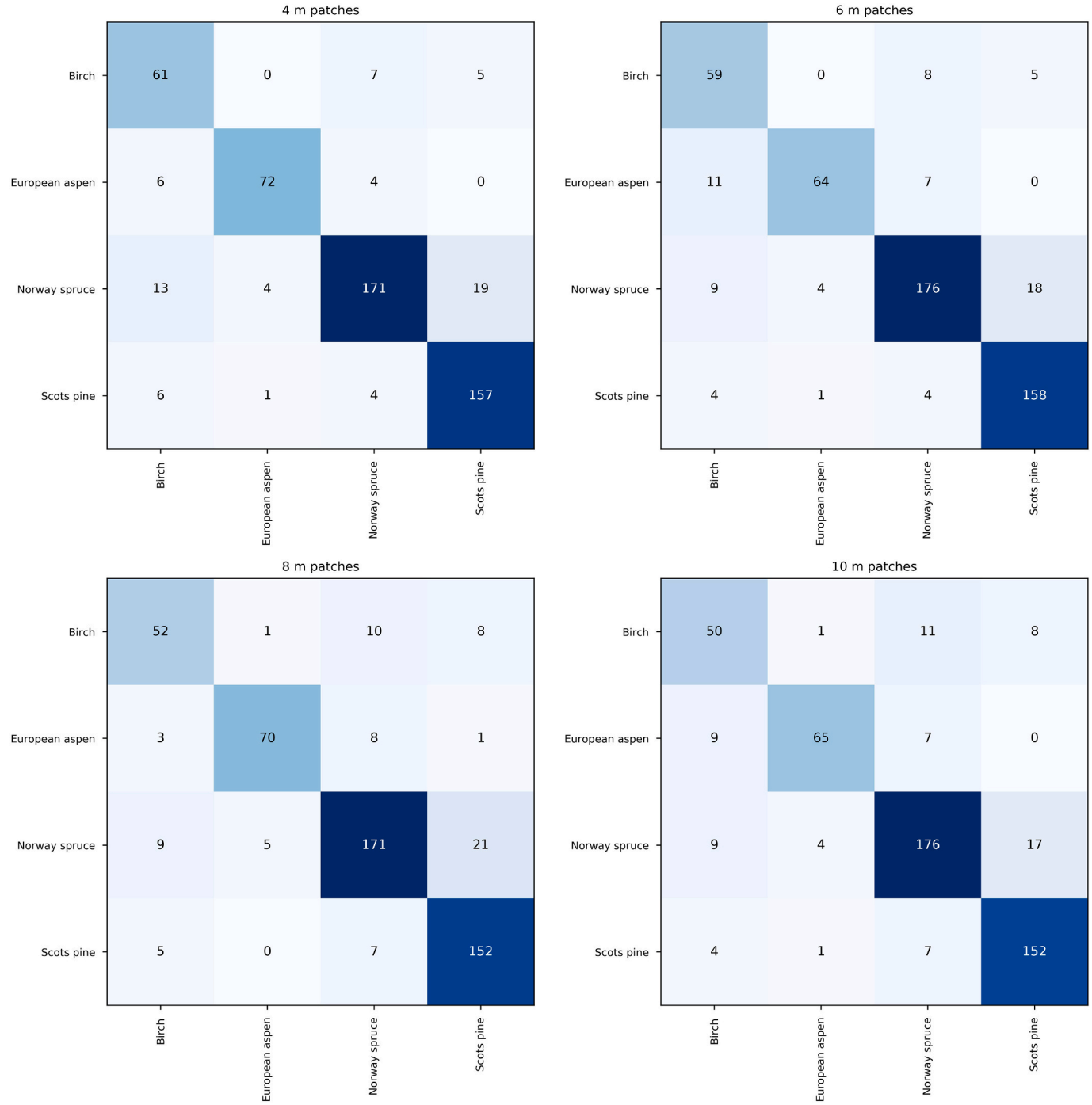


Fig. 7. Confusion matrices for 3D-CNN models. Rows indicate correct labels and columns indicate predicted labels.

of aspen classifications may be due to aspens being the thickest and highest tree species in the area, with wide canopies, which probably resulted in more uniform data than for the other tree species. The wide canopies of aspen decreased the possibility of no non-aspen pixels within the segmented canopy area. Moreover, almost all labeled aspens were measured individually with the RTK-GNSS device resulting in a more accurate positioning.

There are several potential sources of uncertainty in our work. As seen in Fig. 3, there were inconsistencies in both our field data and airborne imagery. The inconsistencies with field data were due to measurement conditions. For field plots, only the plot centers were measured and trees within the plot were located related to the plot center, which may have led to incorrect locations for trees. Individual trees were measured from the central stem position, but the corresponding treetop might have been located elsewhere in our aerial imagery due to for example curvature of the trunk or wind. It is also possible that there were trees with DBH $<$ 150 mm located within a labeled segment which may have affected to the spectral signature. While detecting these situations was possible within 9 m field plots, it was practically impossible for individual trees and the edges of the field plots. We addressed these situations by using the segmentation algorithm to capture only the immediate treetop area.

In the case of airborne imagery, the mismatch between different data sources is a well-known challenge in data fusion, due to, for example, differences in spatial resolution or collection time. As seen in Fig. 3, some delineated canopies and treetops seem to be displaced based on

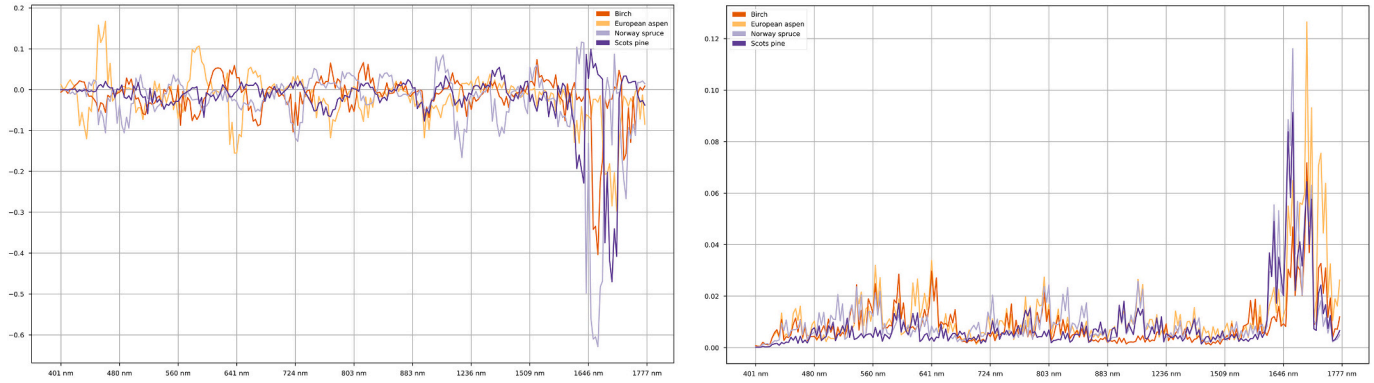


Fig. 8. Left: Average change in prediction confidence when masking spectral wavelengths. Right: Average magnitude of gradients for vanilla backpropagation.

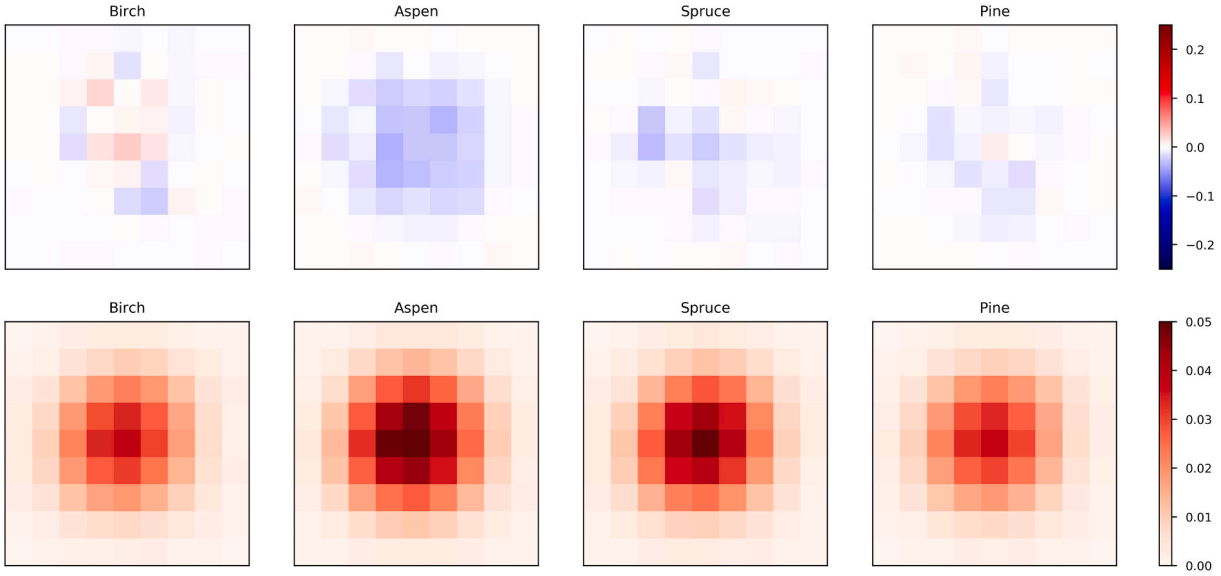


Fig. 9. Top: Average change of prediction confidence when occluding spatial pixels from input data. Bottom: Saliency maps showing average magnitude of gradients.

Table 8
Classification results for the full study area, for all detected trees with maximum height 10 m or more.

Species	Percentage
Scots pine	48.28%
Norway spruce	28.50%
Birch	21.81%
European aspen	1.41%

hyperspectral data, but were in correct locations in the CHM. These kinds of problems were more common in our data at the edges of flight lines, and were most likely artefacts from the orthorectification process. Finally, our models were only able to classify any tree crown into one of the four species and thus omitted other, rarer species present in the area.

The collection method had a large impact on the tree matching rate for ground reference data. The larger the proportion of individually measured trees, the better the detection rate was for each species. On average, individually measured trees had a significantly larger DBH compared to field plot measurements in both unmatched and matched data. However, this could be due to bias in data collection. Individually measured trees were all handpicked by researchers and generally larger than trees on average, whereas the field plot measurements contained all trees within the field plot. Still, for European aspen there was no difference in size between the labeled trees from the field and handpicked data, which may be due to the large average size of the aspen trees in the study area. On the other hand, individually measured trees were positioned more accurately than trees positioned based on the field plot center, which may have eased the process of matching them with the hyperspectral and LiDAR data in the segmentation process.

Our tree matching rate of 42.8% is comparable to or slightly better than the rate in studies with similar LiDAR or CHM data. In previous studies, the achieved detection rates have been from 32% (Dalponte and Coomes, 2016; Kandare et al., 2016) to around 50% (Nevalainen et al., 2017) and up to 63.4% (Hamraz et al., 2019). Comparing these results, however, is not straightforward, because of different field and airborne data collection methods and different limits in minimum height and DBH used in these studies. In this study, we performed tree detection from the CHM instead of LiDAR point clouds. Since our methods utilized spectral features extracted from the upper canopy layer, we wanted to classify the trees clearly visible from birds-eye view and therefore set limitations for both maximum height and DBH. While this limits the applicability of our methods to mature forests, in our study the main point of interest was mature and old-growth trees, especially large and elderly aspen. As for the subject of improving the tree matching rate, algorithms utilizing point cloud data are shown to be more accurate than CHM-level methods, especially in dense, heterogenous canopies (Kandare et al., 2016). However, they require more processing power and are much slower especially for high-density point clouds (Pirotti et al., 2017), and they are prone to oversegmentation (Liu et al., 2019).

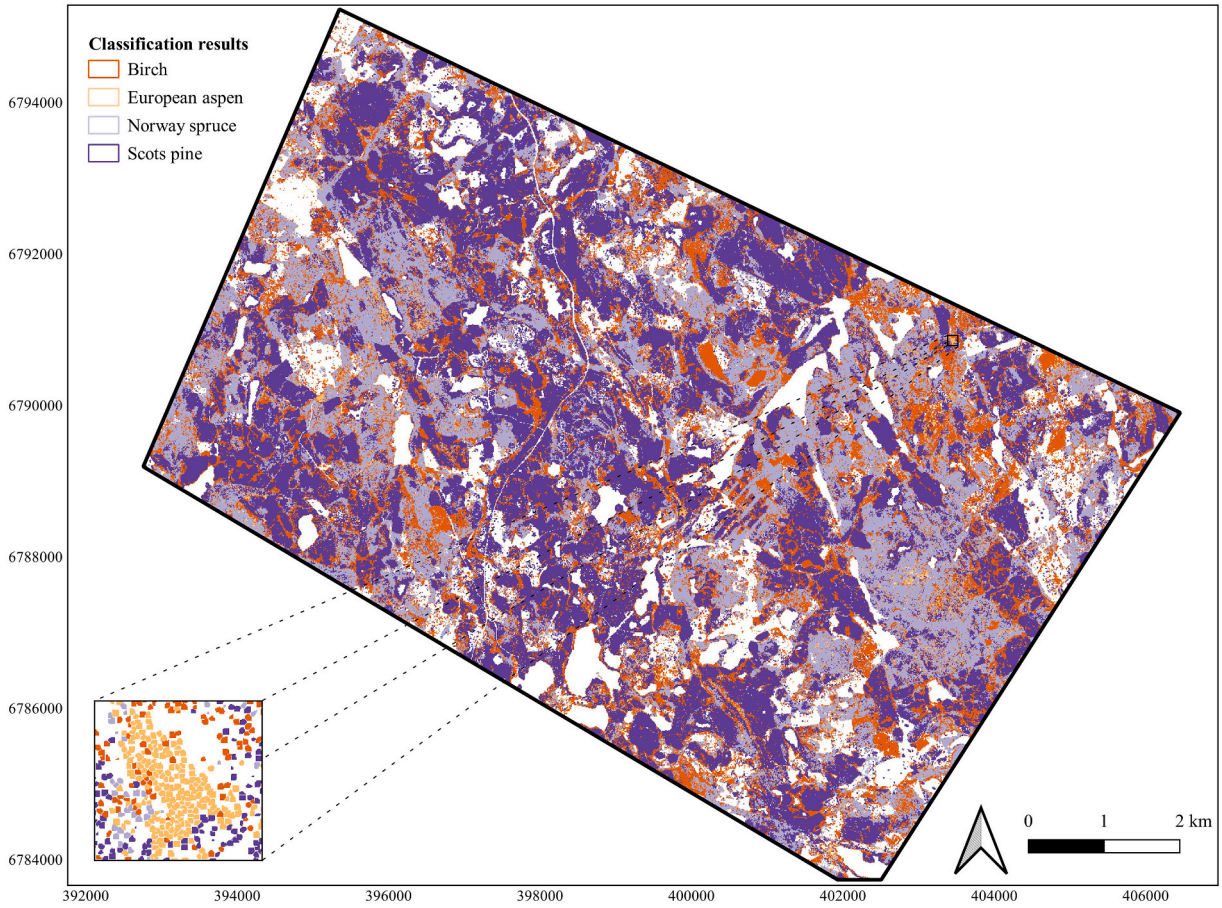


Fig. 10. Wall-to-wall tree species map produced with the best performing 3D-CNN of trees with maximum height of 10 m or more, resampled to 10 m spatial resolution. Lower left: 100 × 100 m window of a larger aspen stand, showing individual trees.

Utilizing more LiDAR parameters or LiDAR derived features either for tree detection or for classification could be an interesting research direction as a continuation of this study. For instance, Hamraz et al. (2019) acquired promising results in separating coniferous and deciduous trees only from LiDAR derived features, and Pölönen et al. (2018) used a normalized CHM as an additional input channel in addition to hyperspectral data in their work. Especially for species that occur sporadically or infrequently in the landscape, imaging spectroscopy as such may not be enough for precise classification (Waser et al., 2014; Roth et al., 2015b). Structural information from LiDAR data in combination with imaging spectroscopy has been found to improve the classification or calibration accuracy (Dalponte et al., 2012).

The OA of 87% in this study corresponds to or outcompetes the OAs acquired by using discriminant analysis methods utilizing hyperspectral data in temperate forest sites (Roth et al., 2015a). In general, boreal forests in Finland have a low number of main tree species which simplifies the identification task. The four species examined in this study dominate more than 97% of the growing stock volume of forests in Finland (Lier et al., 2017). Thus, the classification problem cannot be compared to, e.g., tropical forests with thousands of tree species coexisting in the canopy (Asner et al., 2015). Generally, the more species or classes there are to predict, the larger the field comparison data that is needed (Feret and Asner, 2013; Heinzel and Koch, 2011; Yu et al., 2014).

Based on both occlusion and saliency graphs, the wavelength range of 1650–1700 nm influenced the species classification strongly. In this range, Kokaly and Skidmore (2015) have attributed a narrow spectral absorption feature at 1660 nm to phenolic compounds, whereas tree species with a high lignin content, such as pine and spruce, may have a shift to 1670 nm. For *Populus tremuloides*, phenolic compounds such as condensed tannins and salicinoids have been predicted based on leaf reflectance spectra with many of the important wavelengths in the 1650–1700 nm range (Couture et al., 2016). Spectral features of 1660, 1890 and 2000 nm have been shown to match spectral features of cellulose and lignin polymers and to differ among tree species with different cellulose and lignin content (Buitrago et al., 2018). Viinikka et al. (2020) applied both RF and SVM for the same data used in this work to identify the most important spectral features for different species, and the most significant features for aspen included 1684–1706 nm which coincides with the results of this study by CNN. In this study, the spectral range of 640–660 nm was the most influential VNIR region for deciduous trees and was particularly important in the classification of birch. It co-occurs with the absorbance maxima of about 642 and 660 nm of chlorophyll *b* and a, respectively (Lichtenthaler and Buschmann, 2001). This is in contrast to most studies, which have red edge as the most influential spectral range in VNIR for tree species classification (Heikkinen et al., 2010; Pant et al., 2013), as was also found by Viinikka et al. (2020).

Even though 3D-CNNs were able to achieve better results than other methods, our neural network architecture is most likely not optimal. For most image classification tasks, it is advisable to use an established architecture and pretrained weights as a baseline and only fine-tune the final layer of the model for the task in hand. This process, known as *transfer learning*, has been widely used in remote sensing tasks when the input data is RGB images or even synthetic aperture radar images. However, the relatively low number of labeled samples for deep learning, the variation between band numbers between different datasets, and the usage of different sensors for public benchmarks make transfer learning practically impossible to utilize in hyperspectral image

recognition tasks (Audebert et al., 2019; Paoletti et al., 2019). Because of this, the models in this study were trained from scratch.

One possibility that could be used to improve the classification results is the so-called *self-supervised learning* approach, a subset of unsupervised learning. Unsupervised learning refers to training models without any human-annotated labels, whereas in self-supervised learning, the model is trained with some automatically generated labels. Tasks for self-supervised learning can be, for instance, as simple as predicting whether the original image is rotated (Gidaris et al., 2018), or more advanced, like a clustering-based task (Caron et al., 2018) or a generative task (e.g., Zhu et al. (2017)). The motivation behind self-supervised learning is that by solving tasks with automatically generated labels, the model learns to extract relevant features that can then be fine-tuned for the final task, thus solving the previously mentioned transfer learning problem (Jing and Tian, 2019). Self-supervised learning has been used both in image and video recognition and natural language processing tasks, and utilizing it for hyperspectral imaging would be an interesting research direction.

Using 3D-CNN, we produced a wall-to-wall tree species map for the study area by first segmenting the trees and classifying each tree separately. However, there are already several proposed deep learning approaches to semantic segmentation, such as U-Net (Ronneberger et al., 2015), and one possible direction for future research would be to test their viability for hyperspectral data using segmentation maps produced here as the ground truth. This kind of approach would simplify the required preprocessing steps, as tree delineation or treetop detection would not be needed anymore, and areas with trees lower than a user-defined minimum height could also be labeled. After all, real-world applications, such as biomass or timber yield calculations, have no practical use for individually segmented and classified trees but rather require segmented tree stands consisting of single tree species. Segmentation networks have already been used for multispectral satellite data (see, e.g., Stoian et al. (2019)), but their viability for true hyperspectral data still remains an open question.

Because high resolution hyperspectral data is expensive to collect for large areas using airborne platforms, there is a need to test the feasibility of accurate tree species classification from other sources, such as multispectral satellite images. Currently, there are several satellite instruments in orbit offering multispectral data with high revisit time (e.g., Sentinel-2, Landsat 8). For some species, upscaling with multispectral data seems like a promising opportunity, but there are a number of species for which a more detailed spectral resolution is the only way to improve species recognition. The biodiversity monitoring community is eagerly waiting for open access, high-quality satellite hyperspectral data to support the monitoring of taxonomic and functional diversity (Jetz et al., 2016). There are already hyperspectral satellites in orbit (such as GaoFen-5, PRISMA) and missions in preparation (such as EnMap, CHIME), but data from these are not yet available.

Although valuable for fine-scale planning and investigations, hyperspectral airborne data are typically too detailed for operative land use planning and analysis. Hence, we look forward to linking highly detailed, tree-level airborne data with large-scale spaceborne data. Roth et al. (2015b) showed that upscaling is a viable option in combining hyperspectral datasets of different resolutions. However, linking tree-level data with coarser data raises new questions that require closer inspection, e.g., how to preserve and generalize the spectral information from species that occur in a scattered way and in low numbers. As the spatial resolution is lowered with spaceborne data, the ability to distinguish spectral characteristics of individual tree crowns weakens. This especially affects the non-dominant species. Also, the machine learning-based methods used in this study require extensive training sets, which brings up the challenge of collecting enough field data. Here, we propose that high resolution airborne data could be used for compiling training data for coarser resolution spaceborne data. Hence, the pixel-level variation in spaceborne data could be explained with extensive tree-level data. In the two-phase system, limited field data could be used for training and validating the interpretation methods based on airborne data. Airborne instruments could then be used to collect extensive tree-level data for training the landscape-level species detection utilizing spaceborne data. More accurate landscape-level information on the occurrence of scarce but ecologically significant tree species would benefit ecological modeling (Mononen et al., 2018). Still, a challenge remains on how to collect extensive training data for scarce tree species, such as European aspen. A similar need applies to other species of high ecological importance, such as oak (*Quercus*) and beech (*Fagus*), which support a high number of red-listed invertebrates (Jonsell et al., 1998) in temperate and boreal ecosystems. This study contributes to the knowledge of the modeling capabilities of combined spectral and LiDAR techniques for such scarce species with ecological importance.

6. Conclusions

In this work, we presented a workflow for tree species classification from high resolution hyperspectral and LiDAR imagery, from accurately matching ground reference data with airborne imagery to producing wall-to-wall classification maps. In addition, we compared the performance of random forest, a support vector machine, a gradient boosting machine, a feedforward neural network and a convolutional neural network for this task. The study shows that species identification can be conducted with a high accuracy with the given methods and RS data. Even though the implemented CNN models were most likely not optimal, they were nevertheless able to outperform RF, GBM, SVM and ANN. We aim to use the tree species maps produced in this study as training data for larger scale classification tasks. The development of these kinds of methods is crucial when operationalizing big remote sensing data in biodiversity and ecosystem monitoring.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the Integrated Biodiversity Conservation and Carbon Sequestration in the Changing Environment (IBC-Carbon) (project number 312559), the Strategic Research Council, the Academy of Finland; e-shape has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement 820852 and Mapping and Assessment for Integrated ecosystem Accounting (MAIA) under grant agreement 817527; 2019-FI-ENVECO (Eurostat Grants 2019): Finnish Ecosystem Observatory project funded by the Ministry of the Environment.

The authors wish to acknowledge the CSC – IT Center for Science, Finland, for generous computational resources and excellent user support. Also, the authors would like to thank the editors and reviewers for their thorough comments and suggestions which greatly helped us to improve the final manuscript.

References

Sothe, C., De Almeida, C.M., Schimalski, M.B., La Rosa, L.E., Castro, J.D., Feitosa, R.Q., Dalponte, M., Lima, C.L., Liesenberg, V., Miyoshi, G.T., Tommaselli, A.M., 2020. Comparative performance of convolutional neural network, weighted and conventional support vector machine and random forest for classifying tree species using hyperspectral and photogrammetric data. GISci. Remote Sens. 57, 369–394. https://doi.org/10.1080/15481603.2020.1712102.

Asner, G.P., Martin, R.E., 2009. Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests. Front. Ecol. Environ. 7, 269–276. https://doi.org/10.1890/070152.

Asner, G.P., Martin, R.E., Anderson, C.B., Knapp, D.E., 2015. Quantifying forest canopy traits: imaging spectroscopy versus field survey. Remote Sens. Environ. 158, 15–27. https://doi.org/10.1016/j.rse.2014.11.011.

Audebert, N., Le Saux, B., Lefevre, S., 2019. Deep learning for classification of hyperspectral data: a comparative review. IEEE Geosci. Remote Sens. Mag. 7, 159–173. https://doi.org/10.1109/MGRS.2019.2912563.

Baldeck, C.A., Asner, G.P., Martin, R.E., Anderson, C.B., Knapp, D.E., Kellner, J.R., Wright, S.J., 2015. Operational tree species mapping in a diverse tropical forest with airborne imaging spectroscopy. PLoS One 10, e0118403. https://doi.org/10.1371/journal.pone.0118403.

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13, 281–305. http://scikit-learn.sourceforge.net.

Buitrago, M.F., Skidmore, A.K., Groen, T.A., Hecker, C.A., 2018. Connecting infrared spectra with plant traits to identify species. ISPRS J. Photogramm. Remote Sens. 139, 183–200. https://doi.org/10.1016/j.isprsjprs.2018.03.013.

Caron, M., Bojanowski, P., Joulin, A., Douze, M., 2018. Deep clustering for unsupervised learning of visual features. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 139–156. https://doi.org/10.1007/978-3-030-01264-9_9.

Couture, J.J., Singh, A., Rubert-Nason, K.F., Serbin, S.P., Lindroth, R.L., Townsend, P.A., 2016. Spectroscopic determination of ecologically relevant plant secondary metabolites. Methods Ecol. Evol. 7, 1402–1412. https://doi.org/10.1111/2041-210X.12596.

Dalponte, M., Coomes, D.A., 2016. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. Methods Ecol. Evol. 7, 1236–1245. https://doi.org/10.1111/2041-210X.12575.

Dalponte, M., Bruzzone, L., Gianelle, D., 2012. Tree species classification in the southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. Remote Sens. Environ. 123, 258–270. https://doi.org/10.1016/j.rse.2012.03.013.

Dalponte, M., Ørka, H.O., Ene, L.T., Gobakken, T., Næsset, E., 2014. Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. Remote Sens. Environ. 140, 306–317. https://doi.org/10.1016/j.rse.2013.09.006.

Dalponte, M., Frizzera, L., Gianelle, D., 2019. Individual tree crown delineation and tree species classification with hyperspectral and LiDAR data. PeerJ 2019. https://doi.org/10.7717/peerj.6227.

Deng, J., Dong, W., Socher, R., Li, L.J., Kai, L., Li, F.F., 2009. ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Institute of Electrical and Electronics Engineers (IEEE), pp. 248–255. https://doi.org/10.1109/cvpr.2009.5206848.

Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2016.08.013.

Feret, J.B., Asner, G.P., 2013. Tree species discrimination in tropical forests using airborne imaging spectroscopy. IEEE Trans. Geosci. Remote Sens. 51, 73–84. https://doi.org/10.1109/TGRS.2012.2199323.

Gidaris, S., Singh, P., Komodakis, N., 2018. Unsupervised representation learning by predicting image rotations. In: 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, International Conference on Learning Representations. ICLR. http://arxiv.org/abs/1803.07728.

Goetz, A.F., Vane, G., Solomon, J.E., Rock, B.N., 1985. Imaging spectrometry for earth remote sensing. Science 228, 1147–1153. https://www.researchgate.net/publication/6105820. https://doi.org/10.1126/science.228.4704.1147.

Hamraz, H., Jacobs, N.B., Contreras, M.A., Clark, C.H., 2019. Deep learning for conifer/deciduous classification of airborne LiDAR 3D point clouds representing individual trees. ISPRS J. Photogramm. Remote Sens. 158, 219–230. https://doi.org/10.1016/j.isprsjprs.2019.10.011.

Heikkinen, V., Tokola, T., Parkkinen, J., Korpela, I., Jääskeläinen, T., 2010. Simulated multispectral imagery for tree species classification using support vector machines. IEEE Trans. Geosci. Remote Sens. 48, 1355–1364. https://doi.org/10.1109/TGRS.2009.2032239.

Heinzel, J., Koch, B., 2011. Exploring full-waveform LiDAR parameters for tree species classification. Int. J. Appl. Earth Obs. Geoinf. 13, 152–160. https://doi.org/10.1016/j.jag.2010.09.010.

Howard, J., Gugger, S., 2020. Fastai: A Layered API for Deep Learning, vol. 11. Information 2020, pp. 108–111. https://doi.org/10.3390/INFO11020108.

CSC – IT Center for Science Finland, 2020. Supercomputer Puhti Is Now Available for Researchers - Supercomputer Puhti Is Now Available for Researchers. CSC Company Site. https://www.csc.fi/en/-/supertietokone-puhti-on-avattu-tutkijoiden-kayttoon.

Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: 32nd International Conference on Machine Learning, ICML 2015. International Machine Learning Society (IMLS), pp. 448–456.

Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F.W., Asner, G.P., Guralnick, R., Kattge, J., Latimer, A.M., Moorcroft, P., Schaepman, M.E., Schildhauer, M.P., Schneider, F.D., Schrodt, F., Stahl, U., Ustin, S.L., 2016. Monitoring plant functional diversity from space. Nat. Plant. https://doi.org/10.1038/NPLANTS.2016.24.

Jing, L., Tian, Y., 2019. Self-supervised Visual Feature Learning with Deep Neural Networks: A Survey. http://arxiv.org/abs/1902.06162.

Jones, T.G., Coops, N.C., Sharma, T., 2010. Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific northwest, Canada. Remote Sens. Environ. 114, 2841–2852. https://doi.org/10.1016/j.rse.2010.07.002.

Jonsell, M., Weslien, J., Ehnström, B., 1998. Substrate requirements of red-listed saproxylic invertebrates in Sweden. Biodivers. Conserv. 7, 749–764. https://doi.org/10.1023/A:1008888319031.

Kandare, K., Ørka, H.O., Chan, J.C.W., Dalponte, M., 2016. Effects of forest structure and airborne laser scanning point cloud density on 3D delineation of individual tree crowns. Eur. J. Remote Sens. 49, 337–359. https://doi.org/10.5721/EuJRS20164919.

Kandare, K., Ørka, H.O., Dalponte, M., Næsset, E., Gobakken, T., 2017. Individual tree crown approach for predicting site index in boreal forests using airborne laser scanning and hyperspectral data. Int. J. Appl. Earth Obs. Geoinf. 60, 72–82. https://doi.org/10.1016/j.jag.2017.04.008.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems, pp. 3147–3155. https://github.com/Microsoft/LightGBM.

Kivinen, S., Koivisto, E., Keski-Saari, S., Poikolainen, L., Tanhuanpää, T., Kuzmin, A., Viinikka, A., Heikkinen, R.K., Pykälä, J., Virkkala, R., Vihervaara, P., Kumpula, T., 2020. A keystone species, European aspen (*Populus tremula* L.), in boreal forests: ecological role, knowledge needs and mapping using remote sensing. For. Ecol. Manag. https://doi.org/10.1016/j.foreco.2020.118008.

Kokaly, R.F., Skidmore, A.K., 2015. Plant phenolics and absorption features in vegetation reflectance spectra near 1.66 μm. Int. J. Appl. Earth Obs. Geoinf. 43, 55–83. https://doi.org/10.1016/j.jag.2015.01.010.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105. https://doi.org/10.1145/3065386.

Latva-Karjanmaa, T., Penttilä, R., Siitonen, J., 2007. The demographic structure of European aspen (Populus tremula) populations in managed and old-growth boreal forests in eastern Finland. Can. J. For. Res. 37, 1070–1081. http://www.nrcresearchpress.com/doi/10.1139/X06-289. https://doi.org/10.1139/X06-289.

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L. D., 1989. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1, 541–551. https://doi.org/10.1162/neco.1989.1.4.541.

LeCun, Y.A., Bottou, L., Orr, G.B., Müller, K.R., 2012. Efficient BackProp. In: Neural Networks: Tricks of the Trade. Springer, Berlin, Heidelberg, pp. 9–48. https://link.springer.com/chapter/10.1007/978-3-642-35289-8_3. https://doi.org/10.1007/978-3-642-35289-8_3.

Lichtenthaler, H.K., Buschmann, C., 2001. Chlorophylls and carotenoids: measurement and characterization by UV-VIS spectroscopy. Curr. Protocol. Food Anal. Chem. 1, F4.3.1–F4.3.8. https://doi.org/10.1002/0471142913.faf0403s01.

Lier, M., Korhonen, K., Tuomainen, T., Viitanen, J., Mutanen, A., 2017. Finland's forests 2017. In: Based on FOREST EUROPE Criteria and Indicators of Sustainable Forest Management. Technical Report. Natural Resources Institute Finland, Luke. http://urn.fi/URN:NBN:fi-fe2019091628400.

Liu, L., Lim, S., Shen, X., Yebra, M., 2019. A hybrid method for segmenting individual trees from airborne lidar data. Comput. Electron. Agric. 163, 104871. https://doi.org/10.1016/j.compag.2019.104871.

Loshchilov, I., Hutter, F., 2019. Decoupled weight decay regularization. In: ICLR, p. 2019.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: a meta-analysis and review. ISPRS J. Photogramm. Remote Sens. https://doi.org/10.1016/j.isprsjprs.2019.04.015.

Maltamo, M., Packalen, P., 2014. Species-Specific Management Inventory in Finland. In: Forestry Applications of Airborne Laser Scanning. Springer, pp. 241–252. https://doi.org/10.1007/978-94-017-8663-8_12.

Maltamo, M., Pesonen, A., Korhonen, L., Kouki, J., Vehmas, M., Eerikäinen, K., 2015. Inventory of aspen trees in spruce dominated stands in conservation area. For. Ecosys. 2, 12. https://doi.org/10.1186/s40663-015-0037-4.

Maschler, J., Atzberger, C., Immitzer, M., 2018. Individual tree crown segmentation and classification of 13 tree species using Airborne hyperspectral data. Remote Sens. 10, 1218. http://www.mdpi.com/2072-4292/10/8/1218. https://doi.org/10.3390/rs10081218.

Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans. Geosci. Remote Sens. 42, 1778–1790. https://doi.org/10.1109/TGRS.2004.831865.

Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications – moving from data reproduction to spatial prediction. Ecol. Model. 411 https://doi.org/10.1016/j.ecolmodel.2019.108815.

Modzelewska, A., Fassnacht, F.E., Stereńczak, K., 2020. Tree species identification within an extensive forest area with diverse management regimes using airborne hyperspectral data. Int. J. Appl. Earth Obs. Geoinf. 84, 101960. https://doi.org/10.1016/j.jag.2019.101960.

Mononen, L., Auvinen, A.P., Packalen, P., Virkkala, R., Valbuena, R., Bohlin, I., Valkama, J., Vihervaara, P., 2018. Usability of citizen science observations together with airborne laser scanning data in determining the habitat preferences of forest birds. For. Ecol. Manag. 430, 498–508. https://doi.org/10.1016/j.foreco.2018.08.040.

Müller, R., Kornblith, S., Hinton, G., 2019. When does label smoothing help?. In: Advances in Neural Information Processing Systems, pp. 4694–4703. http://arxiv.org/abs/1906.02629.

Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. 80, 88–99. https://doi.org/10.1016/S0034-4257(01)00290-5.

Nagasubramanian, K., Jones, S., Singh, A.K., Sarkar, S., Singh, A., Ganapathysubramanian, B., 2019. Plant disease identification using explainable 3D deep learning on hyperspectral images. Plant Methods 15. https://doi.org/10.1186/s13007-019-0479-8.

Nevalainen, O., Honkavaara, E., Tuominen, S., Viljanen, N., Hakala, T., Yu, X., Hyyppä, J., Saari, H., Pölönen, I., Imai, N.N., Tommaselli, A.M., 2017. Individual tree detection and classification with UAV-Based photogrammetric point clouds and hyperspectral imaging. Remote Sens. 9 https://doi.org/10.3390/rs9030185.

Ozbulak, U., 2019. PyTorch CNN Visualizations. https://github.com/utkuozbulak/pytorch-cnn-visualizations.

Packalén, P., Maltamo, M., 2007. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. Remote Sens. Environ. 109, 328–341. https://doi.org/10.1016/j.rse.2007.01.005.

Pant, P., Heikkinen, V., Hovi, A., Korpela, I., Hauta-Kasari, M., Tokola, T., 2013. Evaluation of simulated bands in airborne optical sensors for tree species identification. Remote Sens. Environ. 138, 27–37. https://doi.org/10.1016/j.rse.2013.07.016.

Paoletti, M.E., Haut, J.M., Plaza, J., Plaza, A., 2019. Deep learning classifiers for hyperspectral imaging: a review. ISPRS J. Photogramm. Remote Sens. https://doi.org/10.1016/j.isprsjprs.2019.09.006.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in PyTorch. In: NIPS 2017 Autodiff Workshop: The Future of Gradient-Based Machine Learning Software and Techniques.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830. https://doi.org/10.5555/1953048.2078195.

Pirotti, F., Kobal, M., Roussel, J.R., 2017. A comparison of tree segmentation methods using very high density airborne laser scanner data. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 285–290. https://doi.org/10.5194/isprs-archives-XLII-2-W7-285-2017.

Pölönen, I., Annala, L., Rahkonen, S., Nevalainen, O., Honkavaara, E., Tuominen, S., Viljanen, N., Hakala, T., 2018. Tree Species Identification Using 3D Spectral Data and 3D Convolutional Neural Network, in: Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing. Institute of Electrical and Electronics Engineers (IEEE), pp. 1–5. https://doi.org/10.1109/WHISPERS.2018.8747253.

Poso, S., 1983. Basic features of forest inventory by compartments. Silva Fennica 17, 313–343. http://hdl.handle.net/10138/15179.

Rassi, P., Hyvärinen, E., Juslén, A., Mannerkoski, I., 2010. The 2010 red list of Finnish species. Ympäristöministeriö & Suomen ympäristökeskus, Helsinki 685.

Richter, R., Schläpfer, D., 2002. Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: atmospheric/topographic correction. Int. J. Remote Sens. 23, 2631–2649. https://doi.org/10.1080/01431160110115834.

Richter, R., Schläpfer, D., 2004. Atmospheric/Topographic Correction for Airborne Imagery. https://www.rese-apps.com/software/download. http://www.atcor.info/pdf/atcor4_manual.pdf.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

Roth, K.L., Roberts, D.A., Dennison, P.E., Alonzo, M., Peterson, S.H., Beland, M., 2015a. Differentiating plant species within and across diverse ecosystems with imaging spectroscopy. Remote Sens. Environ. 167, 135–151. https://doi.org/10.1016/j.rse.2015.05.007.

Roth, K.L., Roberts, D.A., Dennison, P.E., Peterson, S.H., Alonzo, M., 2015b. The impact of spatial resolution on the classification of plant species and functional types within imaging spectrometer data. Remote Sens. Environ. 171, 45–57. https://doi.org/10.1016/j.rse.2015.10.004.

Roussel, J.R., Caspersen, J., Béland, M., Thomas, S., Achim, A., 2017. Removing bias from LiDAR-based estimates of canopy height: accounting for the effects of pulse density and footprint size. Remote Sens. Environ. 198, 1–16. https://doi.org/10.1016/j.rse.2017.05.032.

Simonyan, K., Vedaldi, A., Zisserman, A., 2014. Deep inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, in: 2nd International Conference on Learning Representations, ICLR 2014 - Workshop Track Proceedings, International Conference on Learning Representations. ICLR. http://arxiv.org/abs/1312.6034.

Smith, L.N., 2018. A Disciplined Approach to Neural Network Hyper-Parameters: Part 1 – Learning Rate, Batch Size, Momentum, and Weight Decay (arXiv preprint arXiv:1803.09820). http://arxiv.org/abs/1803.09820.

Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M., 2015. Striving for simplicity: the all convolutional net. In: 3rd International Conference on Learning Representations, ICLR 2015 - Workshop Track Proceedings.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Stoian, A., Poulain, V., Inglada, J., Poughon, V., Derksen, D., 2019. Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems. Remote Sens. 11, 1986. https://www.mdpi.com/2072-4292/11/17/1986. https://doi.org/10.3390/rs11171986.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp. 2818–2826. https://doi.org/10.1109/CVPR.2016.308.

Trier, Ø.D., Salberg, A.B., Kermit, M., Rudjord, Ø., Gobakken, T., Næsset, E., Aarsten, D., 2018. Tree species classification in Norway from airborne hyperspectral and airborne laser scanning data. Eur. J. Remote Sens. 51, 336–351. https://doi.org/10.1080/22797254.2018.1434424.

Viinikka, A., Hurskainen, P., Keski-Saari, S., Kivinen, S., Tanhuanpää, T., Mäyrä, J., Poikolainen, L., Vihervaara, P., Kumpula, T., 2020. Detecting European Aspen (Populus tremula L.) in Boreal Forests Using Airborne Hyperspectral and Airborne Laser Scanning Data. Remote Sens. 12, 2610. https://www.mdpi.com/2072-4292/12/16/2610. https://doi.org/10.3390/rs12162610.

Waser, L.T., Küchler, M., Jütte, K., Stampfer, T., 2014. Evaluating the potential of worldview-2 data to classify tree species and different levels of ash mortality. Remote Sens. 6, 4515–4545. http://www.mdpi.com/2072-4292/6/5/4515. https://doi.org/10.3390/rs6054515.

Yu, L., Liang, L., Wang, J., Zhao, Y., Cheng, Q., Hu, L., Liu, S., Yu, L., Wang, X., Zhu, P., Li, X., Xu, Y., Li, C., Fu, W., Li, X., Li, W., Liu, C., Cong, N., Zhang, H., Sun, F., Bi, X., Xin, Q., Li, D., Yan, D., Zhu, Z., Goodchild, M.F., Gong, P., 2014. Meta-discoveries from a synthesis of satellite-based land-cover mapping research. Int. J. Remote Sens. https://doi.org/10.1080/01431161.2014.930206.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp. 818–833. http://link.springer.com/10.1007/978-3-319-10602-1. https://doi.org/10.1007/978-3-319-10590-1_53.

Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D., 2018. MixUp: beyond empirical risk minimization. In: 6th International Conference on Learning Representations. ICLR 2018 - Conference Track Proceedings. http://arxiv.org/abs/1710.09412.

Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2242–2251. https://github.com/junyanz/CycleGAN. https://doi.org/10.1109/ICCV.2017.244.